

PROYEK AKHIR DATA WRANGLING
ANALISIS FAKTOR SOSIAL EKONOMI TERHADAP KASUS HIV
DAN AIDS DI INDONESIA TAHUN 2024



Disusun Oleh:

Adisti Eka Nabila 24031554119

Cantika Latifatul Nur Ella 24031554023

Dosen Pengampu:

Bu Ulfa Siti Nuraini, S.Stat., M.Stat.

PRODI S1 SAINS DATA
FAKULTAS MATEMATIKA & ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI SURABAYA
2025/2026

DAFTAR ISI

DAFTAR ISI.....	2
BAB I.....	3
PENDAHULUAN.....	3
A. Latar Belakang.....	3
B. Manfaat.....	4
C. Tujuan.....	4
BAB II.....	5
PEMBAHASAN.....	5
A. Teknik Pengambilan Data.....	5
B. Cleaning dan Integrasi.....	6
C. Exploratory Data Analysis (EDA).....	11
D. Data Publishing.....	23
E. Kendala Dan Rencana Tindak Lanjut.....	24
BAB III.....	26
PENUTUP.....	26
Kesimpulan.....	26
Pembagian Kontribusi Anggota.....	27
Lampiran.....	28
Referensi.....	29

BAB I

PENDAHULUAN

A. Latar Belakang

HIV/AIDS masih menjadi salah satu tantangan besar dalam kesehatan masyarakat di Indonesia, dengan jumlah kasus yang terlihat cukup tajam antarprovinsi. Selain dipengaruhi faktor biologis dan perilaku individu, penyebaran HIV juga sangat dipengaruhi oleh kondisi sosial ekonomi masyarakat. UNAIDS (2023) menegaskan bahwa ketidaksetaraan ekonomi, kemiskinan dan keterbatasan akses terhadap layanan kesehatan merupakan faktor struktural yang meningkatkan kerentanan terhadap HIV, terutama pada kelompok usia produktif. Kondisi kemiskinan menyebabkan masyarakat mengalami hambatan dalam mengakses layanan kesehatan, pemeriksaan rutin, serta edukasi pencegahan HIV. Keterbatasan informasi dan akses ini sering kali mengurangi kemampuan individu untuk memperoleh informasi kesehatan yang akurat, sehingga meningkatkan kemungkinan terjadinya perilaku berisiko seperti penggunaan narkoba suntik atau hubungan seksual. Temuan ini diperkuat oleh penelitian dalam Jurnal Bitara oleh Kurniawati (2022) yang menunjukkan bahwa kemiskinan dan rendahnya tingkat pendidikan mendorong tingginya penyebaran HIV di Indonesia.

Selain kemiskinan, tingkat pendidikan juga mempengaruhi tingkat pemahaman masyarakat mengenai HIV/AIDS. Rendahnya pendidikan sering dikaitkan dengan kurangnya kesadaran tentang cara penularan HIV, pentingnya penggunaan alat pelindung diri, serta kesadaran akan pentingnya pemeriksaan kesehatan. Penelitian Mulyanti et al. (2024) mengungkapkan bahwa kesenjangan sosial dalam pendidikan dan status ekonomi sangat mempengaruhi kualitas hidup serta kemampuan pencegahan pada Orang Dengan HIV/AIDS (ODHA). Temuan lain oleh Adiansyah et al. (2023) menegaskan bahwa rendahnya tingkat pendidikan dan status sosial ekonomi juga meningkatkan stigma, yang pada akhirnya memperburuk akses ODHA terhadap layanan kesehatan. Kurniawati (2022) menegaskan bahwa rendahnya tingkat pendidikan dapat memperlemah kemampuan individu dalam melakukan pencegahan, sehingga meningkatkan risiko penularan HIV. Selain itu, faktor pengangguran turut memperburuk situasi, Kementerian Kesehatan RI (2024) menjelaskan bahwa pengangguran dapat memicu tekanan sosial ekonomi yang berpotensi mendorong perilaku berisiko seperti hubungan seksual tidak aman atau penggunaan narkoba suntik. Hal ini sejalan dengan temuan Nugroho et al. (2023) yang menyatakan bahwa sosial ekonomi merupakan determinan penting yang mempengaruhi penyebaran HIV/AIDS di Indonesia, terutama di wilayah dengan tingkat pengangguran tinggi.

Secara lebih luas, faktor sosial ekonomi dapat pula mempengaruhi kondisi sosial masyarakat yang mendorong timbulnya aktivitas berisiko. Menurut Wijaya (2025), tekanan ekonomi dapat mendorong praktik pekerjaan berisiko seperti pekerja seks, yang secara langsung meningkatkan potensi penularan HIV. Berdasarkan berbagai kondisi

tersebut, semakin jelas bahwa kemiskinan, rendahnya pendidikan, dan tingginya pengangguran merupakan salah satu faktor penting yang meningkatkan kerentanan masyarakat terhadap HIV/AIDS. Oleh karena itu, proyek ini berjudul “*Analisis Faktor Sosial Ekonomi terhadap Kasus HIV/AIDS di Indonesia Tahun 2024*”, dengan tujuan untuk memahami bagaimana faktor-faktor sosial ekonomi tersebut berhubungan dengan angka kasus HIV/AIDS di Indonesia.

B. Manfaat

1. Memberikan bukti empiris mengenai hubungan faktor sosial ekonomi dengan kasus HIV/AIDS per provinsi.
2. Menyediakan informasi untuk mengidentifikasi provinsi yang memiliki kerentanan sosial ekonomi tinggi terhadap HIV/AIDS.
3. Menjadi dasar pertimbangan dalam merancang strategi pencegahan dan penanggulangan HIV/AIDS yang memperhatikan kondisi sosial ekonomi masyarakat.

C. Tujuan

1. Menganalisis hubungan faktor sosial ekonomi terhadap kasus HIV/AIDS per provinsi.
2. Mengidentifikasi provinsi dengan risiko sosial ekonomi tinggi yang berkaitan dengan prevalensi HIV/AIDS.
3. Memberikan gambaran bagaimana kondisi sosial ekonomi dapat berpengaruh terhadap penyebaran HIV/AIDS di Indonesia.

BAB II

PEMBAHASAN

A. Teknik Pengambilan Data

Data diperoleh dari beberapa sumber berbeda, yaitu PDF Profil Kesehatan Indonesia 2024 resmi dari Kementerian Kesehatan (Kemenkes), PDF Profil Kemiskinan di Indonesia September 2024 dari Badan Pusat Statistik (BPS), CSV Rata rata lama sekolah tahun 2024 dari BPS, dan CSV Tingkat Pengangguran Terbuka (TPT) berdasarkan provinsi di indonesia dari Open Data Jabar.

1. Scrapping pdf menggunakan *fitz (PyMuPDF)* pada data hiv/aids dari Profil Kesehatan Indonesia 2024.

```
doc1 = fitz.open(pdf_hiv_aids)
page = doc1[3].get_text()
print(page)
```

Kode diatas digunakan untuk membuka dan membaca isi file pdf untuk data hiv/aids, lalu mengambil teks pada halaman ke-4 dikarenakan indeks dimulai dari 0.

2. Scrapping tabel menggunakan *tabula-py* pada data kasus HIV dan AIDS dari dokumen Kemenkes yang memiliki border pada tabelnya.

kode scraping tabel hiv:

```
tables = tabula.read_pdf(
    pdf_hiv_aids,
    pages=472,
    multiple_tables=True,
    lattice=True
)
scrape_hiv_df = tables[0]
```

Membaca tabel dari pdf dengan tabula, yg dimana hanya mengambil tabel pada halaman 472 dan tabel pertama pada halaman tersebut disimpan sebagai DataFrame `scrape_hiv_df`.

kode scraping tabel aids:

```
tables = tabula.read_pdf(
    pdf_hiv_aids,
```

```

        pages=475,
        multiple_tables=True,
        lattice=True
    )
    scrape_aids_df = tables[0]

```

Membaca tabel dari pdf dengan tabula, yg dimana hanya mengambil tabel pada halaman 475 dan tabel pertama pada halaman tersebut disimpan sebagai DataFrame `scrape_hiv_df`.

3. Scrapping pdf menggunakan *pdfplumber* pada data penduduk miskin.

scraping tabel data kemiskinan:

```

pdf_kemiskinan = '/content/drive/Shared drives/Project
Data Wrangling/Profile Kemiskinan di Indonesia.pdf'
doc2 = fitz.open(pdf_kemiskinan)
doc2
data = []
with pdfplumber.open(pdf_kemiskinan) as pdf:
    page = pdf.pages[9]
    table = page.extract_table()
    df = pd.DataFrame(table[1:], columns=table[0])
    data.append(df)
scrap_kemiskinan_df = pd.concat(data, ignore_index=True)

```

Membuka file pdf dan membaca tabel dari pdf dengan pdfplumber dan mengubah tabel pdf menjadi list of lists (baris & kolom), yg dimana hanya mengambil tabel pada halaman 9. Lalu menggabungkan tabel ke satu DataFrame `scrape_kemiskinan_df`.

4. Menyimpan tabel atau data mentah hasil scraping pdf

```

scrape_hiv_df.to_csv("scrape_hiv_raw.csv", index=False)
scrape_aids_df.to_csv("scrape_aids_raw.csv",
index=False)
scrape_kemiskinan_df.to_csv("scrape_kemiskinan_raw.csv",
index=False)

```

B. Cleaning dan Integrasi

1. Data HIV

- a. Membersihkan kolom dan baris


```
df1 = scrape_hiv_df
```

```
df1.columns = ["No", "Provinsi", "2022", "2023",
               "2024"]
df1 = df1.iloc[2:]
df1 = df1.iloc[:-1]
df1 = df1[["Provinsi", "2024"]]
df1.rename(columns={"2024": "Kasus HIV 2024"},
           inplace=True)
```

b. Membersihkan format angka

```
df1["Kasus HIV 2024"] = df1["Kasus HIV
2024"].str.replace(".", "",
                    regex=False).astype(int)
```

c. Merapikan index dan kolom provinsi

```
df1_hiv = df1.reset_index(drop=True)
df1_hiv['Provinsi'] =
df1_hiv['Provinsi'].str.upper()
```

d. Mengecek missing value

```
df1_hiv.info()
df1_hiv.isna().sum()
```

Tahap pembersihan data dilakukan dengan menghapus baris yang tidak relevan, memilih hanya kolom yang diperlukan, menstandarkan nama kolom, serta membersihkan format angka tahun 2024 dengan menghilangkan tanda titik dan mengonversinya menjadi tipe integer. Index tabel dirapikan dan nama provinsi diubah menjadi huruf kapital agar konsisten. Setelah proses cleaning, integritas data dicek menggunakan `info()` untuk memastikan struktur dan tipe data sudah benar, serta `isna().sum()` untuk memastikan tidak terdapat nilai yang hilang sebelum data digunakan untuk analisis.

2. Data AIDS

a. Membersihkan kolom dan baris

```
df2 = scrape_aids_df
df2.columns = ["No", "Provinsi", "2022", "2023",
               "2024", "Kasus_Kumulatif"]
df2 = df2.iloc[2:]
df2 = df2.iloc[:-1]
df2 = df2[["Provinsi", "2024"]]
df2.rename(columns={"2024": "Kasus AIDS 2024"},
           inplace=True)
```

b. Membersihkan format angka

```
df2["Kasus AIDS 2024"] = df2["Kasus AIDS  
2024"].str.replace(".", "",  
regex=False).astype(int)
```

c. Merapikan index dan kolom provinsi

```
df2_aids = df2.reset_index(drop=True)  
df2_aids['Provinsi'] =  
df2_aids['Provinsi'].str.upper()
```

d. Mengecek missing value

```
df2_aids.info()  
df2_aids.isna().sum()
```

Tahap pembersihan data dilakukan dengan menghapus baris yang tidak relevan, memilih hanya kolom yang diperlukan, menstandarkan nama kolom, serta membersihkan format angka tahun 2024 dengan menghilangkan tanda titik dan mengonversinya menjadi tipe integer. Index tabel dirapikan dan nama provinsi diubah menjadi huruf kapital agar konsisten. Setelah proses cleaning, integritas data dicek menggunakan info() untuk memastikan struktur dan tipe data sudah benar, serta isna().sum() untuk memastikan tidak terdapat nilai yang hilang sebelum data digunakan untuk analisis.

3. Data Kemiskinan

```
df3 = scrape_kemiskinan_df  
missing=df3[df3["Provinsi"].isna()]  
df3 = df3.drop(index=[0,1,2]).reset_index(drop=True)  
df3 = df3.iloc[:-1].reset_index(drop=True)  
df3.columns = ["No", "Provinsi", "Perkotaan Mar'24",  
"perkotaan Sep'24", "Perdesaan Mar'24", "Perdesaan  
sep'24", "Total Mar'24", "Jumlah penduduk miskin"]  
df3_kemiskinan = df3[["Provinsi", "Jumlah penduduk  
miskin"]]  
df3_kemiskinan= df3_kemiskinan.replace("\n", " ",  
regex=True)  
df3_kemiskinan.head(7)  
df3_kemiskinan["Jumlah penduduk miskin"] = (  
df3_kemiskinan["Jumlah penduduk miskin"]  
.astype(str)  
.str.replace('.', '', regex=False)  
.str.replace(',', '. ', regex=False)  
.pipe(pd.to_numeric, errors='coerce')  
)
```



```
df3_kemiskinan["Provinsi"] =
df3_kemiskinan["Provinsi"].str.upper().str.strip()
df3_kemiskinan.isna().sum()
```

Proses pengolahan data dilakukan dengan menghapus baris tidak relevan, menyesuaikan nama kolom, serta memilih variabel yang diperlukan untuk analisis. Data kemudian dibersihkan dengan menghilangkan karakter seperti newline, memperbaiki format angka dengan menghapus pemisah ribuan dan mengonversinya menjadi tipe numerik, serta menstandarkan nama provinsi agar konsisten. Setelah semua langkah cleaning dilakukan, pengecekan integritas data dilakukan melalui pemeriksaan adanya nilai hilang untuk memastikan dataset siap digunakan pada tahap analisis berikutnya.

4. Data Pengangguran

```
csv_pengangguran = '/content/drive/Shared drives/Project
Data Wrangling/Tingkat Pengangguran 2024.csv'
df4 = pd.read_csv(csv_pengangguran)
df4_pengangguran = df4[["nama_provinsi", "indeks_tpt"]]
df4_pengangguran.columns = ["Provinsi", "Indeks TPT"]
df4_pengangguran.info()
df4_pengangguran.isna().sum()
```

Pada dataset pengangguran, proses pengolahan dilakukan dengan memilih hanya kolom yang relevan, yaitu nama provinsi dan indeks tingkat pengangguran terbuka, kemudian menormalkan nama kolom agar lebih konsisten dan mudah digunakan. Setelah itu dilakukan pengecekan struktur data menggunakan info() dan pemeriksaan nilai hilang dengan isna().sum() untuk memastikan tidak ada data yang bermasalah sebelum masuk ke tahap analisis lebih lanjut

5. Data Rata” Lama Sekolah

```
csv_rata_lama_sekolah = '/content/drive/Shared
drives/Project Data Wrangling/Rata Rata Lama Sekolah
2024.csv'
df5 = pd.read_csv(csv_rata_lama_sekolah)
df5.info()
missing=df5[df5["Provinsi/Kabupaten/Kota"].isna()]
df5 =
df5.dropna(subset=["Provinsi/Kabupaten/Kota"]).reset_ind
ex(drop=True)
df5 = df5.iloc[:-1].reset_index(drop=True)
```

```

df5["Provinsi/Kabupaten/Kota"] =
df5["Provinsi/Kabupaten/Kota"].str.replace("D I
YOGYAKARTA", "DI YOGYAKARTA")

df5["Provinsi/Kabupaten/Kota"] =
df5["Provinsi/Kabupaten/Kota"].str.replace("KEP. BANGKA
BELITUNG", "KEPULAUAN BANGKA BELITUNG")
df5.columns = ["Provinsi", "Rata-rata Lama Sekolah"]
df5["Rata-rata Lama Sekolah"] =
pd.to_numeric(df5["Rata-rata Lama Sekolah"],
errors="coerce")
df5_sekolah = df5[["Provinsi", "Rata-rata Lama
Sekolah"]]
df5_sekolah =
df5_sekolah[df5_sekolah["Provinsi"].str.isupper()].reset
_index(drop=True)
df5_sekolah["Provinsi"] =
df5_sekolah["Provinsi"].str.upper().str.strip()
df5_sekolah.isna().sum()

```

Dataset rata-rata lama sekolah dibersihkan dengan menghapus baris yang memiliki nilai provinsi kosong, menghilangkan baris footer yang tidak relevan, serta menormalkan nama provinsi yang tidak konsisten. Selanjutnya, nama kolom disederhanakan agar mudah dibaca, nilai rata-rata lama sekolah dikonversi menjadi format numerik, dan hanya baris dengan nama provinsi berhuruf kapital yang dipertahankan untuk memastikan keseragaman. Tahap akhir dilakukan pengecekan nilai hilang untuk memastikan dataset sudah bersih dan siap digunakan dalam analisis.

6. Integrasi Besar

```

df_merge = (
df3_kemiskinan
.merge(df5_sekolah, on="Provinsi", how="outer")
.merge(df4_pengangguran, on="Provinsi", how="outer")
.merge(df1_hiv, on="Provinsi", how="outer")
.merge(df2_aids, on="Provinsi", how="outer"))

```

Tahap integrasi dilakukan dengan menggabungkan lima dataset berdasarkan nama provinsi menggunakan metode *outer*

join, sehingga seluruh provinsi tetap tercakup meskipun ada data yang kosong. Hasil penggabungan ini menghasilkan satu tabel komprehensif yang memuat seluruh indikator—kemiskinan, pendidikan, pengangguran, HIV, dan AIDS yang kemudian siap digunakan pada tahap analisis selanjutnya.

C. Exploratory Data Analysis (EDA)

Tahap *Exploratory Data Analysis* (EDA) dilakukan untuk memahami struktur data, kualitas data, dan pola awal sebelum memasuki tahap pemodelan. Proses eksplorasi meliputi pengecekan kelengkapan data, konsistensi tipe data, statistik deskriptif, dan identifikasi outlier.

1. Pengecekan Missing Value

Pengecekan ini dilakukan untuk memastikan jika seluruh data tidak ada nilai kosong (*missing value*). Berikut kode pengecekan missing value:

```
df_merge.isna().sum()
```

Output:

	0
Provinsi	0
Jumlah penduduk miskin	0
Rata-rata Lama Sekolah	0
Indeks TPT	0
Kasus HIV 2024	0
Kasus AIDS 2024	0

dtype: int64

Berdasarkan hasil pengecekan missing value diatas, menunjukkan bahwa seluruh kolom memiliki nilai, sehingga tidak perlu melakukan proses imputasi nilai yang hilang. Hal ini dilakukan untuk memastikan bahwa analisis dapat dilakukan tanpa risiko bias akibat data yang tidak lengkap.

2. Pengecekan Data Duplicate

Pemeriksaan duplikasi data dilakukan untuk memastikan bahwa tidak ada provinsi yang tercatat lebih dari satu kali. Berikut kode pengecekan data duplicate:

```
df_merge.duplicated().sum()
```

Output:

```
np.int64(0)
```

Berdasarkan hasil pemeriksaan duplikasi data, menunjukkan bahwa tidak ada baris yang memiliki duplikasi data. Artinya setiap observasi mewakili satu provinsi, sehingga dataset sudah sesuai untuk eksplorasi dan analisis lebih lanjut.

3. Pengecekan Tipe Data

Setiap kolom diperiksa untuk memastikan tipe data telah sesuai dengan perhitungan yang diperlukan. Berikut kode pengecekan tipe data:

```
df_merge.dtypes
```

Output:

```
0
Provinsi      object
Jumlah penduduk miskin  float64
Rata-rata Lama Sekolah  float64
Indeks TPT      float64
Kasus HIV 2024    int64
Kasus AIDS 2024   int64
dtype: object
```

Berdasarkan hasil pengecekan tipe data, kolom numerik seperti Jumlah Penduduk Miskin, Rata-rata Lama Sekolah, Indeks TPT, Kasus HIV 2024, Kasus AIDS 2024, dan Rasio HIV/AIDS telah tersimpan dalam format numerik (integer dan float). Sementara itu, Provinsi tersimpan dalam format string (object). Seluruh tipe data sudah konsisten sehingga tidak memerlukan proses konversi tambahan.

4. Statistik Deskriptif

Statistik deskriptif dihitung untuk melihat gambaran umum persebaran tiap variabel. Ringkasan statistik seperti nilai minimum, maksimum, mean, median, hingga standar deviasi membantu mengidentifikasi variasi antarprovinsi. Berikut kode untuk melihat statistik deskriptif:

```
df_merge.describe()
```

Output:

	Jumlah penduduk miskin	Rata-rata Lama Sekolah	Indeks TPT	Kasus HIV 2024	Kasus AIDS 2024
count	38.000000	38.000000	38.000000	38.000000	38.000000
mean	633.018947	8.841316	4.379737	1676.500000	566.736842
std	944.382913	1.227811	1.414222	2394.915208	830.786917
min	41.110000	4.210000	1.320000	134.000000	0.000000
25%	163.310000	8.300000	3.217500	503.500000	143.500000
50%	299.630000	8.885000	4.190000	865.000000	258.500000
75%	698.477500	9.515000	5.712500	1359.000000	457.000000
max	3893.820000	11.490000	6.750000	10460.000000	3713.000000

5. Identifikasi Outlier

Outlier dianalisis menggunakan metode Interquartile Range (IQR) pada setiap variabel numerik. Outlier terdeteksi pada beberapa provinsi, terutama pada variabel Jumlah Penduduk Miskin, Kasus HIV 2024, Kasus AIDS 2024 & Rasio HIV/AIDS. Berikut kode identifikasi outlier:

```
df6 = df_merge
num_cols = [
    "Jumlah penduduk miskin",
    "Rata-rata Lama Sekolah",
    "Indeks TPT",
    "Kasus HIV 2024",
    "Kasus AIDS 2024",
    "Rasio HIV AIDS"
]

df_num = df6[num_cols]
Q1 = df_num.quantile(0.25)
Q3 = df_num.quantile(0.75)
IQR = Q3 - Q1

lower = Q1 - 1.5 * IQR
upper = Q3 + 1.5 * IQR
outlier_mask = (df_num < lower) | (df_num > upper)
outliers = df6[outlier_mask.any(axis=1)]

print("Outliers detected (per provinsi)")
print(outliers)

print("\nOutlier Boundaries (per variabel)")
print(pd.DataFrame({
    "Lower Bound": lower,
    "Upper Bound": upper
}))
```

Output:

```

Outliers detected (per provinsi)
Provinsi Jumlah penduduk miskin Rata-rata Lama Sekolah \
5 DKI JAKARTA 449.07 11.49
8 JAWA BARAT 3668.35 8.87
9 JAWA TENGAH 3396.34 8.02
10 JAWA TIMUR 3893.82 8.28
19 MALUKU 293.99 10.26
24 PAPUA BARAT 108.28 7.86
26 PAPUA PEGUNUNGAN 331.12 4.21
28 PAPUA TENGAH 287.54 6.12
37 SUMATERA UTARA 1110.92 9.93

Indeks TPT Kasus HIV 2024 Kasus AIDS 2024 Rasio HIV AIDS
5 6.21 5027 1820 2.762088
8 6.75 9893 3239 3.054338
9 4.78 6191 2058 3.008260
10 4.19 10460 3713 2.817129
19 6.11 861 189 4.555556
24 4.13 827 184 4.494565
26 1.32 559 0 0.000000
28 2.75 1311 433 3.027714
37 5.60 3250 1316 2.469605

Outlier Boundaries (per variabel)
Lower Bound Upper Bound
Jumlah penduduk miskin -639.441250 1501.22875
Rata-rata Lama Sekolah 6.477500 11.33750
Indeks TPT -0.525000 9.45500
Kasus HIV 2024 -779.750000 2642.25000
Kasus AIDS 2024 -326.750000 927.25000
Rasio HIV AIDS 1.851131 4.10611

```

```

df_num = df6[[
    "Jumlah penduduk miskin",
    "Rata-rata Lama Sekolah",
    "Indeks TPT",
    "Kasus HIV 2024",
    "Kasus AIDS 2024",
    "Rasio HIV AIDS"
]]

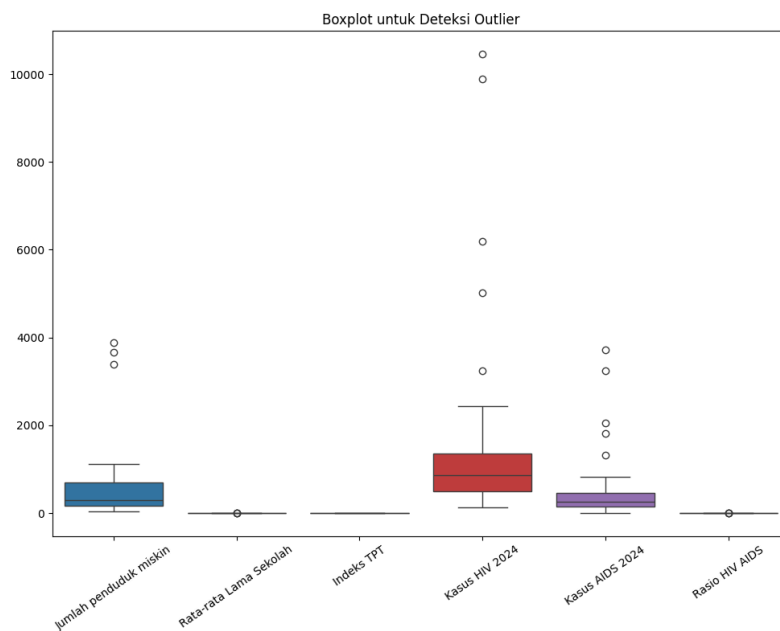
plt.figure(figsize=(10, 8))

sns.boxplot(data=df_num)

plt.title("Boxplot untuk Deteksi Outlier")
plt.xticks(rotation=35)
plt.tight_layout()

plt.show()

```



Berdasarkan hasil pengecekan, outlier ditemukan di beberapa provinsi seperti DKI Jakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Maluku, Papua Barat, Papua Pegunungan, Papua Tengah, dan Sumatera Utara. Batas outlier untuk setiap variabel juga telah dihitung, misalnya jumlah penduduk miskin dianggap outlier jika berada di luar rentang -639.44 hingga 1501.23, dan kasus HIV 2024 dianggap outlier jika berada di luar -779.75 hingga 2642.25. Temuan-temuan ini menunjukkan adanya ketimpangan antar-provinsi yang perlu diperhatikan pada analisis lanjutan.

6. Analisis Rasio HIV/AIDS per Provinsi

Pada tahap ini dilakukan perhitungan rasio antara jumlah kasus HIV dan kasus AIDS untuk setiap provinsi. Rasio ini digunakan untuk melihat seberapa besar proporsi penderita HIV yang telah berkembang menjadi AIDS, sehingga dapat menjadi indikator awal terkait tingkat penanganan kesehatan, atau keterlambatan diagnosis di masing-masing wilayah.

Perhitungan rasio dilakukan dengan membagi jumlah kasus HIV dengan jumlah kasus AIDS per provinsi. Setelah rasio dihitung, data diurutkan dari nilai tertinggi hingga terendah untuk mengidentifikasi provinsi dengan rasio paling menonjol. Berikut kode dan outputnya:

```
rasio_values = []

for hiv, aids in zip(df_merge["Kasus HIV 2024"],
df_merge["Kasus AIDS 2024"]):
    if aids == 0:
        rasio_values.append(0)
    else:
        rasio_values.append(hiv / aids)

df_merge["Rasio HIV AIDS"] = rasio_values
df_merge.sort_values("Rasio HIV AIDS",
ascending=False).head(10)
```

	Provinsi	Jumlah penduduk miskin	Rata-rata Lama Sekolah	Indeks TPT	Kasus HIV 2024	Kasus AIDS 2024	Rasio HIV AIDS
19	MALUKU	293.99	10.26	6.11	861	189	4.55556
24	PAPUA BARAT	108.28	7.86	4.13	827	184	4.49456
4	DI YOGYAKARTA	430.47	9.92	3.48	928	250	3.71200
14	KALIMANTAN TIMUR	211.88	10.02	5.14	1375	389	3.53470
18	LAMPUNG	939.30	8.36	4.19	1220	351	3.47578
2	BANTEN	777.49	9.23	6.68	2434	711	3.42334
7	JAMBI	272.70	8.90	4.48	313	93	3.36591
21	NUSA TENGGARA BARAT	658.60	7.87	2.73	627	189	3.31746
16	KEPULAUAN BANGKA BELITUNG	78.58	8.33	4.63	277	84	3.29761
33	SULAWESI TENGGARA	305.27	9.42	3.09	646	196	3.29591

```

df_sorted = df_merge.sort_values(by="Rasio HIV
AIDS", ascending=False).head(10)

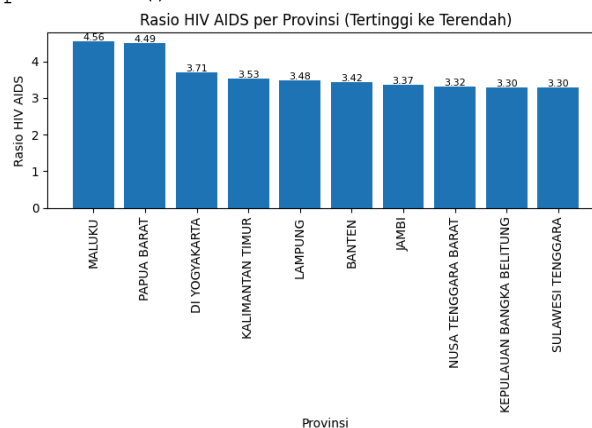
plt.figure(figsize=(7,5))
plt.bar(df_sorted["Provinsi"], df_sorted["Rasio
HIV AIDS"])

plt.xlabel("Provinsi")
plt.ylabel("Rasio HIV AIDS")
plt.title("Rasio HIV AIDS per Provinsi (Tertinggi
ke Terendah)")

for i, v in enumerate(df_sorted["Rasio HIV
AIDS"]):
    plt.text(i, v + 0.025, f"{v:.2f}",
ha='center', fontsize=8)

plt.xticks(rotation=90)
plt.tight_layout()
plt.show()

```

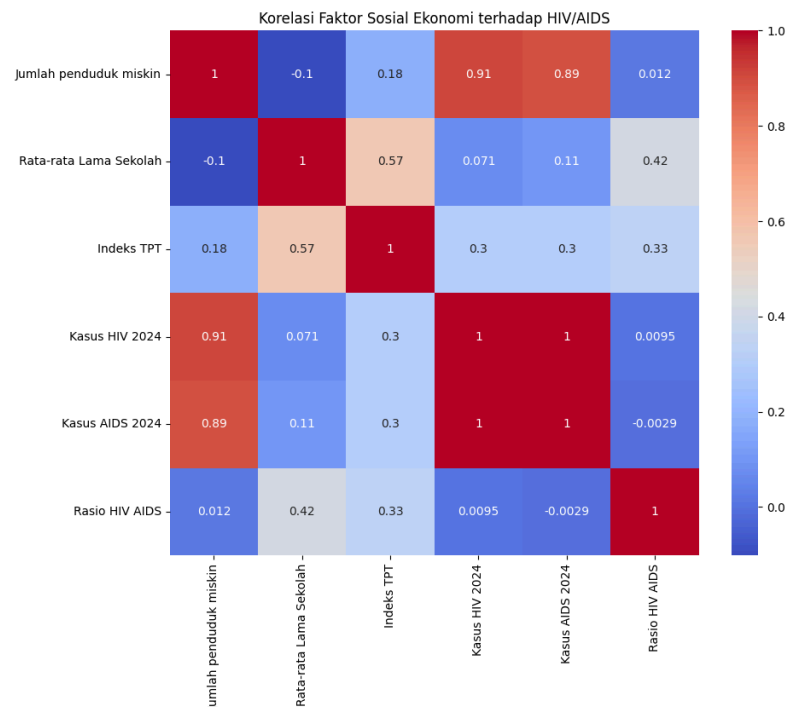


7. Analisis Korelasi Antar Variabel

Analisis korelasi dilakukan menggunakan heatmap untuk melihat kekuatan hubungan antara faktor sosial ekonomi (Jumlah penduduk miskin, Rata-rata lama sekolah, dan indeks TPT) dan kasus HIV & AIDS. Berikut kode heatmap:


```
plt.figure(figsize=(10,8))
sns.heatmap(df_merge.corr(numeric_only=True),
annot=True, cmap="coolwarm")
plt.title("Korelasi Faktor Sosial Ekonomi terhadap
HIV/AIDS")
plt.show()
```

Outputnya:



Visualisasi heatmap korelasi menunjukkan hubungan antar variabel sosial ekonomi dengan kasus HIV dan AIDS pada tahun 2024. Warna merah menandakan korelasi positif yang kuat, sedangkan biru menunjukkan korelasi negatif atau sangat lemah.

a. Jumlah Penduduk Miskin

Variabel jumlah penduduk miskin memiliki korelasi positif yang sangat kuat dengan:

- Kasus HIV 2024 ($r = 0.91$)
- Kasus AIDS 2024 ($r = 0.89$)

Artinya, daerah yang memiliki jumlah penduduk miskin lebih tinggi cenderung menghasilkan jumlah kasus HIV dan AIDS yang lebih besar. Temuan ini menunjukkan bahwa semakin tinggi tingkat kemiskinan suatu wilayah, semakin besar risiko penyebaran HIV/AIDS. Kondisi ini sejalan dengan studi

Adiansyah et al. (2023) yang menemukan bahwa kerentanan sosial ekonomi tidak hanya memengaruhi risiko penularan, tetapi juga meningkatkan stigma terhadap ODHA. Stigma tersebut membuat individu enggan mengakses layanan kesehatan atau melakukan tes secara rutin, sehingga memperburuk penyebaran kasus akibat keterlambatan diagnosis. Selain itu, hubungan dengan variabel sosial lain seperti Rata-rata Lama Sekolah menunjukkan korelasi negatif lemah ($r = -0.10$), yang berarti wilayah dengan kemiskinan tinggi cenderung memiliki tingkat pendidikan yang lebih rendah.

b. Rata-rata Lama Sekolah

Secara umum, rata-rata lama sekolah tidak memiliki hubungan kuat terhadap kasus HIV maupun AIDS:

- Kasus HIV 2024 ($r = 0.071$)
- Kasus AIDS 2024 ($r = 0.11$)

Korelasi yang lemah ini menunjukkan bahwa lama pendidikan formal tidak secara langsung berkaitan dengan tingginya kasus HIV/AIDS. Namun, variabel ini memiliki korelasi cukup moderat dengan Rasio HIV/AIDS ($r = 0.42$), yang dapat mengindikasikan bahwa daerah dengan pendidikan lebih baik memiliki kemampuan deteksi, pelaporan, dan penanganan kasus yang lebih optimal. Hal ini sejalan dengan hasil penelitian Mulyanti & Yulis (2024) yang menegaskan bahwa determinan sosial seperti pendidikan dan dukungan sosial sangat berpengaruh terhadap kualitas hidup ODHA. Pendidikan yang lebih baik meningkatkan pengetahuan tentang pencegahan, mengurangi stigma, dan memperbaiki kemampuan individu mengelola kondisi kesehatannya.

c. Indeks TPT (Tingkat Pengangguran Terbuka)

Indeks pengangguran menunjukkan korelasi positif moderat terhadap:

- Kasus HIV ($r = 0.30$)
- Kasus AIDS ($r = 0.30$)

Pengangguran yang lebih tinggi cenderung selaras dengan meningkatnya kasus HIV/AIDS secara moderat. Hal ini dapat dikaitkan dengan tekanan ekonomi dan sosial yang berdampak pada perilaku atau kondisi yang meningkatkan risiko, seperti migrasi, kurangnya stabilitas sosial, dan keterbatasan akses layanan kesehatan. Indeks TPT juga memiliki hubungan sedang terhadap Rata-rata Lama Sekolah ($r = 0.57$), yPenelitian

Adiansyah et al. (2023) memperkuat hal ini dengan menyatakan bahwa ketidakstabilan sosial ekonomi berkaitan erat dengan persepsi stigma, sehingga ODHA di komunitas dengan kondisi ekonomi rendah lebih rentan terhadap diskriminasi dan kesenjangan akses kesehatan.

d. Hubungan antar Kasus HIV dan AIDS

Nilai korelasi antara Kasus HIV 2024 dan Kasus AIDS 2024 berada pada $r = 1$, menunjukkan hubungan yang sangat kuat dan searah. Hal ini sesuai dengan karakteristik epidemiologi, di mana tingginya kasus HIV biasanya diikuti oleh tingginya jumlah kasus AIDS, jika penanganan tidak optimal. Kondisi ini mempertegas pentingnya faktor sosial ekonomi sebagai akar masalah, karena daerah dengan kemiskinan dan pengangguran tinggi sering memiliki deteksi dini rendah, edukasi kurang, dan stigma kuat, sehingga kasus HIV berkembang menjadi AIDS tanpa intervensi yang memadai.

8. Analisis Scatter Plot (Hubungan Variabel Sosial Ekonomi dengan Kasus HIV/AIDS)

Dalam memahami hubungan langsung antar variabel, scatter plot dengan trendline (regresi plot) digunakan. Visualisasi dibuat untuk tiap kombinasi variabel sosial ekonomi terhadap kasus HIV 2024 & AIDS 2024. Setiap titik pada grafik mewakili satu provinsi, dan angka indeks provinsi disertakan sebagai label agar pola antarwilayah terlihat jelas. Garis regresi membantu menggambarkan arah dan kecenderungan hubungan. Berikut kode scatter plot:

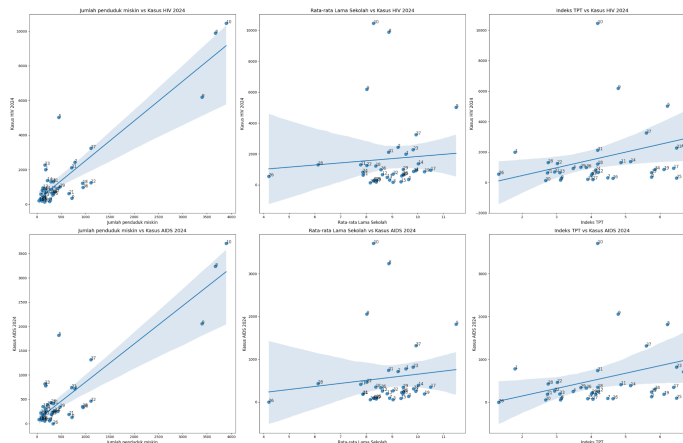
```
fig, axes = plt.subplots(2, 3, figsize=(28, 18))
targets = ["Kasus HIV 2024", "Kasus AIDS 2024"]
features = ["Jumlah penduduk miskin", "Rata-rata Lama Sekolah", "Indeks TPT"]
for row, target in enumerate(targets):
    for col, feature in enumerate(features):
        ax = axes[row, col]
        sns.regplot(
            ax=ax,
            data=df_merge,
            x=feature,
            y=target,
            scatter_kws={"s": 80}
        )
        ax.set_title(f"{feature} vs {target}",
            fontsize=14)
        ax.set_xlabel(feature, fontsize=12)
```

```

ax.set_ylabel(target, fontsize=12)
for i in range(len(df_merge)):
    ax.text(
        df_merge[feature].iloc[i],
        df_merge[target].iloc[i],
        str(df_merge.index[i]),
        fontsize=12,
        alpha=0.8
    )
plt.tight_layout()
plt.show()

```

Output:



Berdasarkan hasil visualisasi data pada tahun 2024, diperoleh gambaran hubungan antara faktor sosial ekonomi dengan jumlah kasus HIV dan AIDS di berbagai wilayah di Indonesia. Variabel yang dianalisis meliputi jumlah penduduk miskin, rata-rata lama sekolah, dan indeks Tingkat Pengangguran Terbuka (TPT).

a. Hubungan Jumlah Penduduk Miskin dengan Kasus HIV dan AIDS

Hasil analisis menunjukkan adanya hubungan positif yang cukup kuat antara jumlah penduduk miskin dengan jumlah kasus HIV maupun AIDS. Hal ini terlihat dari pola sebaran data yang cenderung naik seiring dengan meningkatnya jumlah penduduk miskin serta kemiringan garis regresi yang cukup tajam. Secara sosial, kondisi kemiskinan berpotensi membatasi akses masyarakat terhadap layanan kesehatan, layanan tes HIV, edukasi kesehatan reproduksi, serta pengobatan yang memadai. Selain itu, tekanan ekonomi juga dapat meningkatkan kerentanan terhadap perilaku berisiko, seperti penggunaan narkoba suntik atau aktivitas seksual berisiko tanpa perlindungan. Dengan demikian, temuan ini menunjukkan bahwa faktor ekonomi memiliki peran penting dalam tingginya angka kasus HIV dan AIDS.

- b. Hubungan Rata-rata Lama Sekolah dengan Kasus HIV dan AIDS
Berbeda dengan variabel kemiskinan, hubungan antara rata-rata lama sekolah dengan kasus HIV dan AIDS terlihat lemah dan tidak signifikan secara visual. Pada grafik, sebaran data tampak cukup acak dengan kemiringan garis regresi yang rendah. Hal ini mengindikasikan bahwa tingkat pendidikan formal yang diukur melalui rata-rata lama sekolah belum cukup mampu menjelaskan variasi jumlah kasus HIV/AIDS di suatu wilayah. Ada kemungkinan bahwa meskipun tingkat pendidikan tinggi, pengetahuan spesifik mengenai HIV/AIDS, perilaku pencegahan, dan akses informasi kesehatan tidak serta-merta meningkat. Selain itu, faktor perilaku individu, budaya, stigma sosial, dan kualitas pendidikan kesehatan reproduksi juga memiliki pengaruh yang signifikan tetapi belum tercermin dalam variabel ini.
- c. Hubungan Indeks TPT dengan Kasus HIV dan AIDS
Hasil analisis juga menunjukkan adanya hubungan positif sedang antara indeks TPT dan jumlah kasus HIV maupun AIDS. Wilayah dengan tingkat pengangguran yang lebih tinggi cenderung menunjukkan jumlah kasus yang lebih besar. Pengangguran dapat memicu tekanan psikologis, ketidakstabilan ekonomi, serta meningkatkan kemungkinan individu terlibat dalam aktivitas berisiko, baik secara langsung maupun tidak langsung. Meskipun demikian, tingkat hubungan ini tidak sekuat hubungan dengan kemiskinan, menunjukkan bahwa pengangguran berkontribusi, tetapi bukan faktor utama tunggal.

9. Eksplorasi Variabel Sosial Ekonomi per Provinsi

Eksplorasi variabel sosial ekonomi dilakukan untuk memahami kondisi setiap provinsi. Pada tahap ini faktor sosial ekonomi menjadi indikator utama yang dianalisis dengan mengurutkan nilai tertinggi ke terendah untuk melihat provinsi dengan karakteristik sosial ekonomi yang paling menonjol. Berikut kode dan outputnya:

```
df_merge.sort_values("Jumlah penduduk miskin",  
ascending=False).head(10)
```

	Provinsi	Jumlah penduduk miskin	Rata-rata Lama Sekolah	Indeks TPT	Kasus HIV 2024	Kasus AIDS 2024	Rasio HIV AIDS
10	JAWA TIMUR	3893.82	8.28	4.19	10460	3713	2.817129
8	JAWA BARAT	3668.35	8.87	6.75	9893	3239	3.054338
9	JAWA TENGAH	3396.34	8.02	4.78	6191	2058	3.008260
37	SUMATERA UTARA	1110.92	9.93	5.60	3250	1316	2.469605
22	NUSA TENGGARA TIMUR	1107.94	8.02	3.02	1255	465	2.698925
36	SUMATERA SELATAN	948.84	8.57	3.86	986	332	2.969880
18	LAMPUNG	939.30	8.36	4.19	1220	351	3.475783
2	BANTEN	777.49	9.23	6.68	2434	711	3.423347
0	ACEH	718.96	9.64	5.75	352	130	2.707692
31	SULAWESI SELATAN	711.77	8.86	4.19	2120	741	2.860999

```
df_merge.sort_values("Rata-rata Lama Sekolah",
ascending=True).head(10)
```

	Provinsi	Jumlah penduduk miskin	Rata-rata Lama Sekolah	Indeks TPT	Kasus HIV 2024	Kasus AIDS 2024	Rasio HIV AIDS
26	PAPUA PEGUNUNGAN	331.12	4.21	1.32	559	0	0.000000
28	PAPUA TENGAH	287.54	6.12	2.75	1311	433	3.027714
11	KALIMANTAN BARAT	333.99	7.78	4.86	1309	415	3.154217
24	PAPUA BARAT	108.28	7.86	4.13	827	184	4.494565
21	NUSA TENGGARA BARAT	658.60	7.87	2.73	627	189	3.317460
9	JAWA TENGAH	3396.34	8.02	4.78	6191	2058	3.008260

```
df_merge.sort_values("Indeks TPT",
ascending=False).head(10)
```

	Provinsi	Jumlah penduduk miskin	Rata-rata Lama Sekolah	Indeks TPT	Kasus HIV 2024	Kasus AIDS 2024	Rasio HIV AIDS
8	JAWA BARAT	3668.35	8.87	6.75	9893	3239	3.054338
2	BANTEN	777.49	9.23	6.68	2434	711	3.423347
23	PAPUA	161.07	9.82	6.48	2275	819	2.777778
25	PAPUA BARAT DAYA	96.81	8.39	6.48	291	97	3.000000
17	KEPULAUAN RIAU	124.96	10.50	6.39	949	352	2.696023
5	DKI JAKARTA	449.07	11.49	6.21	5027	1820	2.762088
19	MALUKU	293.99	10.26	6.11	861	189	4.555556
34	SULAWESI UTARA	173.30	9.84	5.85	869	312	2.785256
35	SUMATERA BARAT	315.43	9.44	5.75	654	238	2.747899
0	ACEH	718.96	9.64	5.75	352	130	2.707692

Hasil eksplorasi menunjukkan bahwa provinsi dengan jumlah penduduk miskin terbesar didominasi wilayah berpenduduk besar seperti Jawa Timur, Jawa Tengah, dan Jawa Barat, yang secara alami memiliki beban sosial ekonomi yang lebih berat. Sebaliknya, provinsi dengan

jumlah penduduk miskin paling sedikit cenderung berasal dari wilayah berpenduduk relatif kecil seperti Kalimantan Utara dan Papua Selatan.

Pada variabel Rata-rata Lama Sekolah, terlihat bahwa provinsi di Indonesia bagian timur seperti Papua Pegunungan dan Papua Tengah memiliki angka lama sekolah yang paling rendah, menandakan adanya tantangan akses pendidikan. Sementara provinsi seperti DKI Jakarta, DI Yogyakarta, dan Kepulauan Riau menempati posisi teratas dengan lama sekolah tertinggi.

Untuk Indeks TPT, provinsi berpenduduk padat seperti Banten, Sulawesi Utara, dan Aceh berada pada kelompok dengan tingkat pengangguran lebih tinggi, sedangkan wilayah seperti Papua Pegunungan, Sulawesi Barat, dan Nusa Tenggara Barat memiliki tingkat pengangguran yang relatif rendah.

D. Data Publishing

1. Penyimpanan Data Mentah & Scraping

Seluruh file yang diperoleh dari berbagai sumber Kemenkes, BPS, dan Open Data Jabar disimpan pada folder yang mencakup:

- Profil Kesehatan Indonesia 2024.pdf
- Profile Kemiskinan di Indonesia.pdf
- Rata Rata Lama Sekolah 2024.csv
- Tingkat Pengangguran 2024.csv
- scrape_hiv_raw.csv
- scrape_aids_raw.csv
- scrape_kemiskinan_raw.csv

2. Penyimpanan Data Hasil Cleaning

Setelah proses data cleaning; standardisasi provinsi, konversi numerik, penghapusan simbol ribuan, perbaikan hasil scraping PDF, dan filter kolom, setiap dataset disimpan ulang dalam folder yang mencakup:

- data_hiv_clean.csv
- data_aids_clean.csv
- data_kemiskinan_clean.csv
- data_pengangguran_clean.csv
- data_lama_sekolah_clean.csv

3. Penyimpanan Data Hasil Integrasi

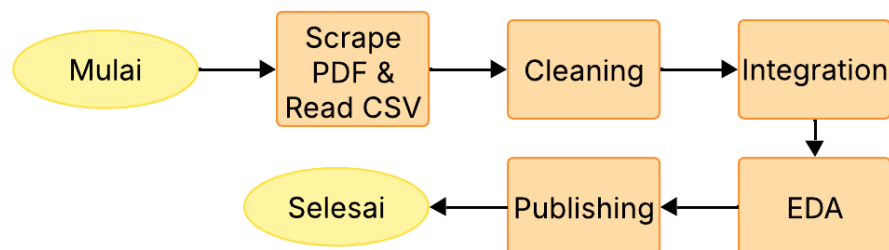
Seluruh dataset digabungkan berdasarkan variabel Provinsi menggunakan merge. Hasil integrasi disimpan pada file:

- DATA FINAL.csv

4. Pipeline

- a. Pengambilan Data

- Mengunduh dataset dari Kementerian Kesehatan, BPS dan Open Data Jabar.
 - Melakukan scraping PDF menggunakan pdfplumber dan tabula
 - Menyimpan seluruh file ke folder
- b. Data Cleaning
- Standarisasi nama provinsi.
 - Menghapus baris yang bukan data, seperti header duplikat.
 - Membersihkan format angka.
 - Konversi kolom numerik dari object → int/float.
- c. Integrasi Data
- Filter/Subset beberapa kolom data.
 - Menggabungkan seluruh dataset berdasarkan kolom Provinsi.
- d. EDA
- Mengecek missing value, duplicate, dan outlier
 - Menampilkan statistik deskriptif seluruh variabel.
 - Membuat visualisasi dasar: histogram, boxplot, scatterplot, dan heatmap.
 - Menyortir provinsi berdasarkan: jumlah penduduk miskin, lama sekolah, TPT, dan rasio HIV/AIDS.



E. Kendala Dan Rencana Tindak Lanjut

1. Kendala Selama Pengerjaan

- a. Kesulitan menemukan dataset yang lengkap untuk 38 provinsi. Tidak semua sumber menyediakan data sosial ekonomi secara menyeluruh bagi 38 provinsi.
- b. Perbedaan format file antar sumber data. Terdapat dua data PDF dan dua data CSV, sehingga file PDF harus di-*scrape* menjadi bentuk tabel dan dikonversi menjadi file CSV agar semua dataset bisa dibaca oleh Pandas.

- c. Struktur tabel dalam file PDF yang tidak selalu rapi. Beberapa tabel PDF memiliki *merge cell*, kolom yang tidak terdeteksi dan layout yang tidak konsisten, akibatnya saat *scrape* sebagian kolom terbaca sebagai NaN.
- d. Format angka yang menyebabkan tipe data tidak konsisten, dalam file penduduk miskin, format angka ditulis seperti, 1.234,56, sehingga angka tersebut dibaca sebagai object, bukan numerik. Sehingga saat cleaning harus mengganti koma menjadi titik desimal, dan menghapus tanda titik pemisah ribuan.
- e. Ketidak konsistenan penamaan provinsi antar dataset. Terdapat beberapa nama provinsi yang tidak konsisten, contoh D I Yogyakarta dengan DI Yogyakarta, Kep. bangka belitung dan Kepulauan bangka belitung.

2. Rencana Tindak Lanjut

Setelah seluruh proses pengumpulan data, pembersihan (data cleansing), serta integrasi empat sumber data berhasil dilakukan, tahap selanjutnya adalah menentukan metode analisis lanjutan yang paling sesuai dengan karakteristik dataset.

- a. Menambah variabel sosial ekonomi lainnya untuk analisis yang lebih tepat dan komprehensif.
- b. Cluster provinsi, karena data setelah wrangling tidak memiliki variabel target atau label khusus yang ingin diprediksi. Clustering memungkinkan pengelompokan provinsi berdasarkan kesamaan karakteristik sosial-ekonomi dan tingkat kasus HIV/AIDS. Melalui pendekatan ini, provinsi dapat terbagi ke dalam beberapa kelompok, cluster dengan tingkat kemiskinan dan kasus HIV/AIDS tinggi, cluster dengan tingkat pendidikan rendah tetapi kasus kesehatan menengah, dan cluster dengan kondisi sosial-ekonomi baik serta angka kasus rendah. Analisis ini bermanfaat untuk mengidentifikasi pola kemiripan antar provinsi serta membantu penentuan prioritas intervensi pada wilayah yang memiliki risiko lebih tinggi.
- c. Model Regresi, metode regresi dapat digunakan apabila penelitian diarahkan untuk membangun model prediktif, seperti memprediksi jumlah kasus HIV, kasus AIDS, atau rasio HIV/AIDS. Karena semua variabel bersifat numerik, model yang dapat diterapkan mencakup regresi linear, multiple linear regression, Random Forest Regressor, atau XGBoost Regressor. Metode regresi juga memungkinkan penilaian kontribusi masing-masing faktor (feature importance) terhadap variabel target. Namun, regresi bersifat opsional dan tidak wajib digunakan apabila fokus penelitian adalah analisis deskriptif.

BAB III PENUTUP

Kesimpulan

Berdasarkan hasil analisis, faktor sosial ekonomi memiliki hubungan yang erat dengan variasi jumlah kasus HIV/AIDS di Indonesia tahun 2024, meskipun dengan tingkat pengaruh yang berbeda pada setiap variabel. Faktor kemiskinan menunjukkan hubungan yang paling kuat, di mana provinsi dengan jumlah penduduk miskin lebih tinggi cenderung memiliki jumlah kasus HIV/AIDS yang lebih besar. Kondisi ini berkaitan dengan keterbatasan akses terhadap layanan kesehatan, pemeriksaan HIV, edukasi kesehatan reproduksi, serta layanan pengobatan yang memadai. Hal ini sejalan dengan temuan UNAIDS (2023) dan Kurniawati (2022) yang menegaskan bahwa kemiskinan merupakan faktor struktural utama yang meningkatkan kerentanan terhadap HIV.

Selain itu, tingkat pengangguran juga menunjukkan keterkaitan dengan peningkatan jumlah kasus HIV/AIDS. Tekanan sosial ekonomi akibat pengangguran dapat meningkatkan risiko individu terlibat dalam perilaku berisiko, seperti hubungan seksual tidak aman dan penggunaan narkoba suntik. Temuan ini sejalan dengan Kementerian Kesehatan RI (2024) dan Nugroho et al. (2023) yang menyatakan bahwa pengangguran merupakan salah satu determinan sosial penting dalam penyebaran HIV/AIDS.

Berbeda dengan variabel kemiskinan dan pengangguran, faktor pendidikan yang diukur melalui rata-rata lama sekolah memiliki hubungan yang lebih lemah. Hal ini menunjukkan bahwa pendidikan formal belum tentu secara langsung berkorelasi dengan rendahnya kasus HIV/AIDS, karena pengaruhnya dipengaruhi pula oleh kualitas edukasi kesehatan, akses informasi, serta faktor sosial budaya lainnya. Temuan ini sejalan dengan Mulyanti et al. (2024) dan Adiansyah et al. (2023) yang menyebutkan bahwa pendidikan lebih berperan secara tidak langsung melalui peningkatan kesadaran dan penurunan stigma.

Secara keseluruhan, hasil penelitian ini menegaskan bahwa upaya penanggulangan HIV/AIDS perlu dilakukan secara komprehensif dengan tidak hanya berfokus pada aspek medis, tetapi juga pada perbaikan kondisi sosial ekonomi, seperti pengentasan kemiskinan, penurunan pengangguran, serta penguatan pendidikan dan literasi kesehatan masyarakat. Pendekatan ini penting untuk menekan angka kasus HIV/AIDS secara lebih berkelanjutan di Indonesia.

Pembagian Kontribusi Anggota

Nama	NIM	Kontribusi
Adisti Eka Nabila	24031554119	<ul style="list-style-type: none">● Pada kode python:<ul style="list-style-type: none">- Scraping PDF untuk kasus kemiskinan- Cleaning-integrasi kasus kemiskinan dan rata-rata lama sekolah- EDA● Laporan:<ul style="list-style-type: none">- Pendahuluan: latar belakang, tujuan, dan manfaat- Teknik pengambilan data- Cleaning dan integrasi● PPT
Cantika Latifatul N. E.	24031554023	<ul style="list-style-type: none">● Pada kode python:<ul style="list-style-type: none">- Scraping PDF untuk kasus HIV & AIDS- Cleaning-integrasi kasus HIV, kasus AIDS, tingkat pengangguran- Integrasi besar- EDA● Laporan:<ul style="list-style-type: none">- Ekplorasi Data Analysis(EDA)- Data Publishing- Kendala & Rencana Tindak Lanjut● PPT

Lampiran

Link Colab:  [KEL 26_DATA WRANGLING.ipynb](#)

Link Data Publishing (GitHub): [GitHub_Kel 26](#)

Referensi

- Adiansyah, M. T., Ramani, A., & Baroya, N. M. (2023). Determinants of Stigma on People Living With Hiv and Aids in Indonesia (Evidence From 2017 Idhs Data). *The Indonesian Journal of Public Health*, 18(2), 291. <https://doi.org/10.20473/ijph.v18i2.2023.291-301>
- Joint United Nations Programme on HIV/AIDS (UNAIDS). (2023). The path that ends AIDS: UNAIDS global AIDS update 2023. UNAIDS. <https://www.unaids.org>
- Kementerian Kesehatan Republik Indonesia. (2024). *Laporan perkembangan HIV/AIDS dan Infeksi Menular Seksual (IMS) di Indonesia*. Direktorat Jenderal Pencegahan dan Pengendalian Penyakit.
- Kurniawati, Y. (2020). Pengaruh tingkat pendidikan dengan kejadian HIV/AIDS. *Jurnal Bitara*, 3(2), 45–52. <https://doi.org/10.30737/jubitar.v3i2.1674>
- Mulyanti, M., & Yulis, D. M. (2024). Analysis of social determinants with quality of life in people with HIV/AIDS at the voluntary counseling test clinic. *Jurnal Ilmiah Kesehatan Sandi Husada*, 13(2), 250-256. <https://doi.org/10.35816/jiskh.v13i2.1211>
- Nugroho, E., Wanti, P. A., Suci, C. W., Raharjo, B. B., & Najib, N. (2023). Social determinants of stunting in Indonesia. *KEMAS: Jurnal Kesehatan Masyarakat*, 18(4), 717–726. <https://doi.org/10.15294/kemas.v18i4.40875>